ELSEVIER International Journal of Approximate Reasoning 18 (1998) 93–106

INTERNATIONAL JOURNAL OF APPROXIMATE REASONING

Simple data filtering in rough set systems

Ivo Düntsch [a,*], Günther Gediga [b,1]

[a] *School of Information and Software Engineering, University of Ulster, Newtownabbey, BT 37 0QB, UK*

[b] *FB Psychologie/Methodenlehre, Universität Osnabrück, 49069 Osnabrück, Germany*

Received 1 November 1996; accepted 10 May 1997

Abstract

In symbolic data analysis, high granularity of information may lead to rules based on a few cases only for which there is no evidence that they are not due to random choice, and thus have a doubtful validity. We suggest a simple way to improve the statistical strength of rules obtained by rough set data analysis by identifying attribute values and investigating the resulting information system. This enables the researcher to reduce the granularity within attributes without assuming external structural information such as probability distributions or fuzzy membership functions. © 1998 Elsevier Science Inc.

Keywords: Rough sets; Data filtering; Statistical validation

1. Introduction

Major aims of data analysis are to discover which features or attributes are relevant for data description and/or prediction, and to filter out the irrelevant ones. A symbolic approach to achieve these aims is Rough Set Data Analysis (RSDA) which has been developed by Pawlak and his co-workers since the early 1980s, and has recently received wider attention as a means of data mining, cf. [1]. The original view behind the rough set model was the observation that:

* Corresponding author. Tel.: +44 1232 368976; e-mail: i.duentsch@ulst.ac.uk.

[1] Equal authorship implied. E-mail: ggediga@luce.psycho.Uni-Osnabrueck.DE.

PII S0888-613X(97)10005-6

> The information about a decision is usually vague because of uncertainty and imprecision coming from many sources... Vagueness may be caused by *granularity* of representation of the information. Granularity may introduce an ambiguity to explanation or prescription based on vague information [2].

The rough set model is intended to be a structural, non-numerical method of information analysis; quantitative aspects are of only secondary interest.

Knowledge representation in the rough set model is done via information systems which are a tabular form of an OBJECT → ATTRIBUTE VALUE relationship. An example which we will use to demonstrate our approach is given in Table 1.

We interpret this information system as follows:

- $x_1, \ldots, x_8$ are persons.
- The attribute m is a combined measure of medical indicators for the risk of a heart attack, while p is a combined measure of psychological indicators, see e.g. [3,4].
- The values of the risk measures are:
  1. NO RISK,
  2. SMALL RISK,
  3. MEDIUM RISK,
  4. HIGH RISK,
  5. VERY HIGH RISK.
- The decision variable H is interpreted as the observation of a heart attack within a predefined time span, and we code

  1–HEART ATTACK 0–NO HEART ATTACK

What is given in Table 1 is often called "raw data". These are unfiltered measurements of attributes within the domain under investigation. However, it can be argued that there is no such thing as observed "raw data": Which attributes are chosen, and which measurements are used, are pragmatic decisions by researchers, how they want to represent the dependencies of real life criteria in the best possible way.

Table 1
The sample information system $\mathscr{I}_1$

U	m	P	H
x_1	1	3	0
x_2	3	2	0
x_3	2	1	0
x_4	3	3	0
x_5	2	4	1
x_6	4	1	1
x_7	1	5	1
x_8	5	4	1

In other words, following [5], we adopt the attitude that:

THE RESEARCHER'S CHOICE OF ATTRIBUTES AND MEASUREMENTS ARE PART OF THE MODEL BUILDING PROCESS AND OF THE DATA ANALYSIS.

RSDA is therefore a conditional data analysis strategy, dependent on the choice of attributes and measurement models. The relevance of the chosen attributes and their measurement can be tested by statistical techniques. In [6] we have developed two simple procedures, both based on randomization techniques, which evaluate the validity of prediction based on the approximation quality of attributes of rough set dependency analysis. These procedures seem to be particularly suitable to the soft computing approach of RSDA as a data mining tool which is *data driven*, and does not require outside information. In particular, it is not assumed that the information system under discussion is a representative sample. The reader is invited to consult [7] or [8] for the background and justification of randomization techniques in these situations.

It follows that – like other types of data analysis – RSDA needs a preprocessing step which results in data which is suitable for further analysis. This preprocessing step should be part of the measurement procedure; although highly desirable in certain situations – for example when a system has an empty core or when the obtained rules are based on a few observations – it seems that it has not yet been addressed as an integral part of rough set analysis.

In this paper we shall show that some of the pragmatic aspects of measurement – say, how to choose an attribute coding to produce valid prediction rules – can be investigated in the context of RSDA. Our main tool will be *binary information systems*; these are data tables in which every column has only two values.

The organization of the paper is as follows: Section 2 will give the formal definitions of information systems and the apparatus needed for rough set dependency analysis; we exhibit some structural properties of binary information systems, and describe a procedure how to associate a binary information system with a general information system. We then apply these instruments to reduce granulation of attribute measurements in order to obtain a higher strength of prediction in terms of the statistical significance test of [6].

The procedures discussed below – e.g. the statistical evaluation of rules and the simple data filtering described in this paper – are implemented in our rough set engine GROBIAN [9] which we have used for the numerical calculations.

2. Knowledge representation

An *information system* is a tuple $\mathscr{I} = \langle U, \Omega, V_q, f_q \rangle_{q\in\Omega}$, where

- U is a finite set of objects,

- Ω is a finite set of attributes (features),
- for each $q \in \Omega$,
  - V_q is a set of attribute values for q,
  - each $f_q: U \to V_q$ is an *information function.*

To avoid trivialities, we assume that each V_q has at least two elements. With each subset Q of Ω we associate an equivalence relation θ_Q on U by setting

$$x \equiv \theta_Q\, y \iff f_q(x) = f_q(y) \quad \text{for all } q \in Q. \tag{1}$$

We shall usually write θ_q instead of $\theta_{\{q\}}$, and, more generally, will identify singletons with the element they contain, if no confusion can arise. The partition associated with θ_Q is denoted by $\mathscr{P}(Q)$.

Given Q, $P \subseteq \Omega$, each class X of θ_Q intersects one or more classes Y_i, $i \leqslant k_X$, of θ_P. This leads to Q, P-rules of the form:

Deterministic: $x \in X \to x \in Y_0$.

Indeterministic: $x \in X \to x \in Y_0 \vee Y_1 \vee \cdots \vee Y_{k_X}$, where $k_X > 0$, and $X \cap Y_i \neq \emptyset$ for $i \leqslant k_X$.

$X \in \mathscr{P}(Q)$ is called *P-deterministic* (or just *deterministic* if P is understood) if it is contained in a class of P. We use $Q \to P$ for the conjunction over all Q, P-rules, and call $Q \to P$ *deterministic* if all Q, P-rules are deterministic. In this case, we write $Q \Rightarrow P$, and we have a set of rules with which we can locally replace P by Q.

The approximation quality $\gamma(Q \to P)$ or a rule $Q \to P$ is defined as

$$\gamma(Q \to P) := \frac{\sum |\{X \in \mathscr{P}(Q):\ X \text{ is } P\text{-deterministic}\}|}{|U|}. \tag{2}$$

This is the cardinality of the positive region of P with respect to Q, cf. [10]. Note that $Q \Rightarrow P$ if and only if $\gamma(Q \to P) = 1$.

Given $P \subseteq \Omega$, of particular interest in rough set dependency theory are those attribute sets Q which are minimal with respect to the property that $Q \Rightarrow P$.

A set Q with this property is called a *rule reduct* [2] of P. If $P = \Omega$, we call Q simply a *reduct*. The *core* of $\mathscr{I}$ – denoted by core($\mathscr{I}$) – the intersection of all reducts of Ω. It is not hard to see, that each $P \subseteq \Omega$ has a rule reduct, though this need not be unique. In rough set theory, the core elements are deemed essential for the knowledge representation, and an empty core indicates a high substitution rate among the attributes. This may be due to incomplete preprocessing of the "raw data" which results in an information system in which the granularity is still too high. We shall discuss this problem in Section 4.

It may be worth to point out that forming of rule reducts is a procedure local to the attribute sets involved. In particular, reducts of Ω or its core only describe how the finest partition of U – induced by the whole system – can be

[2] This differs from the usual definition of reduct, see [10].

obtained by (possibly) fewer features than all of Ω. In algebraic terms, a rule reduct Q of P corresponds to one concrete inclusion in $\{\theta_R: R \subseteq \Omega\}$, and affects only θ_P. The statement "attributes in a reduct can replace the whole attribute set" is not globally true, since only one equation in the whole semilattice of induced equivalence relations in U is addressed.

3. Binary information systems

If the value range V_q of an attribute has exactly two elements, q is called a *binary* attribute; if each attribute is binary, $\mathscr{I}$ is called a *binary information system*.

There is a long standing tradition (for example [11,12]) to distinguish between *symmetric* and *asymmetric* binary attributes. In an asymmetric attribute, the values V_q of an attribute q are based on an indicator function whose value we set as $V_q = \{0, 1\}$: If $f_q(x) = 1$ we have an indicator for the existence of a feature – e.g. the appearance of a symptom or the presence of a certain colour – whereas $f_q(x) = 0$ indicates that we know nothing about x with respect to the attribute q.

Suppose that the researcher has identified the set Ω^D of asymmetric binary attributes (which may be empty), and that $V_q = \{0, 1\}$ for every $q \in \Omega^D$. With $\mathscr{I}$ we associate a binary information system $\mathscr{I}^B = \langle U, \Omega^B, \{0, 1\}, f_q^B \rangle_{q \in \Omega^B}$ as follows.

First, let $\Omega^M := \Omega \setminus \Omega^D$, and then set

$$\Omega^B := \{\langle q, v\rangle : q \in \Omega^M, v \in \operatorname{ran}(f_q)\} \cup \Omega^D,$$

where $\operatorname{ran}(f_q)$ denotes the set of values of f_q.

Note that $\Omega \cap \Omega^B = \Omega^D$. The information functions f_t^B are defined as follows: If $t \in \Omega^D$, then $f_t^B := f_t$. Otherwise, there are $q \in \Omega^M$, $v \in V_q$ such that $t = \langle q, v\rangle$, and we set

$$f_t^B(x) := \begin{cases} 1 & \text{if } f_q(x) = v, \\ 0 & \text{otherwise.} \end{cases}$$

In this case, we say that the binary attribute t *belongs to the class* $f_q^{-1}(v)$ of θ_q. A similar construction which, however, does not distinguish between symmetric and asymmetric binary attributes was given in [13].

Table 2 shows the binarization of the example given in Table 1. For better readability we have written m_i resp. p_i instead of $\langle m, i\rangle$, resp. $\langle p, i\rangle$. Furthermore, because prediction in rough set analysis does not take into account (a) symmetry in the decision attribute(s), we describe H simply by its symmetric version.

In going from $\mathscr{I}$ to $\mathscr{I}^B$, the core is reduced unless $\mathscr{I} = \mathscr{I}^B$, as we shall show below. First, we quote a result from [14]:

Table 2
The binarized system $\mathscr{I}_1^B$

U	m					p					H
	m_1	m_2	m_3	m_4	m_5	p_1	p_2	p_3	p_4	p_5	
x_1	1	0	0	0	0	0	0	1	0	0	0
x_2	0	0	1	0	0	0	1	0	0	0	0
x_3	0	1	0	0	0	1	0	0	0	0	0
x_4	0	0	1	0	0	0	0	1	0	0	0
x_5	0	1	0	0	0	0	0	0	1	0	1
x_6	0	0	0	1	0	1	0	0	0	0	1
x_7	1	0	0	0	0	0	0	0	0	1	1
x_8	0	0	0	0	1	0	0	0	1	0	1

Lemma 1. *Let $\mathscr{I}$ be an information system with attribute set Ω. Then, $p \in core(\mathscr{I})$ if and only if $\theta_\Omega \not\subseteq \theta_{\Omega\setminus\{p\}}$.*

Proposition 2. $core(\mathscr{I}^B) = core(\mathscr{I}) \cap \Omega^D$.

Proof. ($\subseteq$) Let $t \in core(\mathscr{I}^B)$, and assume that $t = \langle q, v\rangle \in \Omega^M$. By Lemma 1 there are $x, y \in U$ such that $f_s^B(x) = f_s^B(y)$ for all $s \in \Omega^B \setminus \{t\}$, and $f_t^B(x) \neq f_t^B(y)$. Assume w.l.o.g. that $f_q(x) = v$, $f_q(y) = w$, and $v \neq w$. Then, $s := \langle q, w\rangle \in \Omega^B \setminus \{t\}$, and hence,

$$f_s^B(x) = f_s^B(y) = 1, \qquad \text{i.e. } f_q(x) = f_q(y) = w,$$

a contradiction; thus, core $(\mathscr{I}^B) \subseteq \Omega^D$.

If $q \in \mathtt{core}(\mathscr{I}^B)$ and $q \notin \mathtt{core}(\mathscr{I})$, then $\theta_\Omega = \theta_{\Omega\setminus\{q\}}$, and it follows from the fact that q is binary that $\theta_{\Omega^B\setminus\{q\}} = \theta_{\Omega^B}$. This contradicts $q \in \mathtt{core}(\mathscr{I}^B)$.

($\supseteq$) Let $q \in \mathtt{core}(\mathscr{I}) \cap \Omega^D$. If $q \notin \mathtt{core}(\mathscr{I}^B)$, then, as above,

$$\theta_\Omega = \theta_{\Omega^B} = \theta_{\Omega^B\setminus\{q\}} = \theta_{\Omega^B\setminus\{q\}},$$

contradicting $q \in \mathtt{core}(\mathscr{I})$. □

4. Granularity analysis using binary information systems

Binarization suggests a way to reduce the number of values of an attribute without loss of dependency information in a specific situation, and at the same time increase the statistical significance of the generated rules. We exemplify our procedure in this section, and then present the theoretical background and the statistical justification in the following sections.

Consider again the information system $\mathscr{I}_1$ given in Table 1. One easily sees that $\theta_{\{m,p\}}$ is the identity relation id_U on U, and therefore we have the rule

$$\{m,p\} \Rightarrow H.$$

On the other hand, the statistical rough set analysis of [6] shows that the chance to obtain this dependency by a random process is close to 100%. This is surprising, because the dependency

High medical and high psychological risk leads to heart attack

is obviously present. But note that the latter statement uses far less information than $\mathscr{I}_1$ gathers: There are only the two risk values {high, not high}. If we recode the risk values accordingly, we obtain the information system $\mathscr{I}_2$ of Table 3.

We still have the dependency $\{m,p\} \Rightarrow H$; the statistical analysis, however, shows that the chance to get the same result by random is about 2.8%. Hence, this dependency can be considered significant. The higher statistical strength of the prediction given in $\mathscr{I}_2$ is due to the fact that the risk groups 1, 2, and 3 are identified, as well as the 4 and 5 risk groups. The differences within these risk groups are neglected, and only the difference between the risk groups remains as a characteristic of the set of prediction attributes $Q = \{m,p\}$. This leads to a duplication of rule instances which influence the statistical strength in a positive way. Before we present the general background, we outline the procedure how to get from $\mathscr{I}_1$ to $\mathscr{I}_2$:

1. Build the binary extension $\mathscr{I}_1^B$ as shown in Table 2.
2. Find the binary attributes m_i, p_j for which

$$(\forall x \in U)[f_{m_i}(x) = 1 \rightarrow f_H(x) = 1], \tag{3}$$

$$(\forall x \in U)[f_{p_j}(x) = 1 \rightarrow f_H(x) = 1], \tag{4}$$

and build their union within m, resp. p in the following way: If, for example, $m_{i_0}, \ldots, m_{i_k}$ satisfy (3), then we define a new binary attribute $m_{i_0} \ldots i_k$ by

Table 3
The recoded example of Table 1

U	m	p	H
x_1	0	0	0
x_2	0	0	0
x_3	0	0	0
x_4	0	0	0
x_5	0	1	1
x_6	1	0	1
x_7	0	1	1
x_8	1	1	1

Table 4
The system $(\mathscr{I}_1^B)^*$

U	m_1	m_2	m_3	m_{45}	p_1	p_2	p_3	p_{45}	H
x_1	1	0	0	0	0	0	1	0	0
x_2	0	0	1	0	0	1	0	0	0
x_3	0	1	0	0	1	0	0	0	0
x_4	0	0	1	0	0	0	1	0	0
x_5	0	1	0	0	0	0	0	1	1
x_6	0	0	0	1	1	0	0	0	1
x_7	1	0	0	0	0	0	0	1	1
x_8	0	0	0	1	0	0	0	1	1

$$f_{m_{i_0}\ldots i_k}(x) = 1 \overset{\text{def}}{\iff} f_{m_j}(x) = 1 \quad \text{for some } j \in \{i_0, \ldots, i_k\}$$
$$\iff \max_{j\in\{i_0,\ldots,i_k\}} f_{m_j}(x) = 1$$

and simultaneously replace $m_{i_0} \ldots m_{i_k}$ by the new attribute $m_{i_0} \ldots i_k$.

Because $\{m_4, m_5\}$ as well as $\{p_4, p_5\}$ exhibit this property, we replace the two attributes $m_4, m_5(p_4, p_5)$ by a new aggregate attribute $m_{45}(p_{45})$ to obtain the binary information system $(\mathscr{I}_1^B)^*$ given in Table 4.

3. Similarly, we find the attributes m_i, p_j for which

$$(\forall x \in U)[f_{m_i}(x) = 1 \to f_H(x) = 0],$$

$$(\forall x \in U)[f_{p_j}(x) = 1 \to f_H(x) = 0]$$

and build their union within m, resp. p. We see that only $\{p_2, p_3\}$ has this property, so that we obtain as follows: If H has more classes, we have to perform such a step of collecting positive classes of the independent attributes for each class of H.

4. Perform a rough set dependency analysis with the attributes of this system with respect to the decision attribute H. This results in the system of Table 5.
5. Choose all rule reducts with the smallest cardinality. In the example there is only one collection of attributes that meets this condition, namely, the set $\{m_{45}, p_{45}\}$ (see Table 6).

Because all other binary attributes are superfluous to express the dependency of H from m and p, we finally obtain $\mathscr{I}_2$ of Table 3.

5. The general case

Let $\mathscr{I} = \langle U, \Omega, V_q, f_q\rangle_{q\in\Omega}$ be an information system, where we assume for simplicity that $\operatorname{ran} f_q = V_q$ for all $q \in \Omega$. Suppose that $\Omega = \{q_0, \ldots, q_n\}$ with $V_{q_i} = \{u_i^0, \ldots, u_i^{t(i)}\}$, and let q_i^k be the binary attribute belonging to u_i^k, i.e.

Table 5
Filtered binary system

U	m_1	m_2	m_3	m_{45}	p_1	p_{23}	p_{45}	H
x_1	1	0	0	0	0	1	0	0
x_2	0	0	1	0	0	1	0	0
x_3	0	1	0	0	1	0	0	0
x_4	0	0	1	0	0	1	0	0
x_5	0	1	0	0	0	0	1	1
x_6	0	0	0	1	1	0	0	1
x_7	1	0	0	0	0	0	1	1
x_8	0	0	0	1	0	0	1	1

Table 6
Filtered binary reducts

1	$\{m_1,m_2,m_3,p_{45}\}$
2	$\{m_2,p_1,p_{45}\}$
3	$\{m_2,p_{23},p_{45}\}$
4	$\{m_2,p_1,p_{23}\}$
5	$\{m_{45},p_1,p_{23}\}$
6	$\{m_{45},p_{45}\}$

$$f_{q_i^k}(x) = \begin{cases} 1 & \text{if } f_{q_i}(x) = u_i^k, \\ 0 & \text{otherwise.} \end{cases}$$

It is not hard to see that

$$\theta_{q_i} = \bigcap_{k \leqslant t(i)} \theta_{q_i^k}.$$

Let d be a decision attribute generating the partition $\mathscr{P} = \{P_0, \ldots, P_m\}$ of U. Note that we are not restricting ourselves to a binary decision attribute, but allow an arbitrary partition (which, for example, can be equal to θ_Ω).

For all $i \leqslant n$, $j \leqslant m$ we set

$$M(i,j) := \{k \leqslant t(i) : f_{q_i}^{-1}(u_i^k) \subseteq P_j\}.$$

If $M(i,j) \neq \emptyset$, we define a new binary attribute $W(q_i, P_j)$ by

$$\begin{aligned} f_{W(q_i,P_j)}(x) = 1 &\overset{\text{def}}{\iff} f_{q_i}(x) = u_i^k \quad \text{for some } k \in M(i,j) \\ &\iff \sum_{k \in M(i,j)} f_{q_i^k}(x) = 1. \end{aligned}$$

For all $i \leqslant n$, $j \leqslant m$ we now simultaneously replace the binary attributes belonging to u_i^k, $k \in M(i,j)$, of q_i – i.e. those that correspond to the classes of θ_{q_i} which are wholly contained in P_j – by the single attribute $W(q_i, P_j)$; this corresponds to steps 2, 3 above.

Let $\mathscr{I}_1^B$ be the resulting binary information system. A *local binary rule reduct of* $\mathscr{P}$ now is a rule reduct of $\mathscr{P}$ in the system $\mathscr{I}_1^B$ (Step 4 above). We say "local" rule reduct, since we collect binary attributes within the attributes of the original system. If we collect across attributes as well, we may speak of a *global binary rule reduct*.

Since in going from $\mathscr{I}^B$ to $\mathscr{I}_1^B$ we do not change the number of occurrences of 1's in the rows of $\mathscr{I}^B$, we can replace the attribute values $\{u_i^0, \ldots, u_i^{t(i)}\}$ belonging to q_i by the set

$$\{u_i^k : k \notin \bigcup_{j \leqslant m} M(i,j)\} \cup \{W(q_i, P_j) : M(i,j) \neq \emptyset\}$$

and obtain a new attribute q_i^* by using the obvious attribute function. The classes of the partition associated with q_i^* in the new system are unions of classes of θ_{q_i} in such a way, that elements are identified whose θ_{q_i}-classes are contained in the same element of $\mathscr{P}$.

For $Q \subseteq \Omega$, we let $Q^* = \{q^* : q \in Q\}$. For later use, we show the following lemma.

Lemma 3. *If $Q \subseteq \Omega$, and Q^*, d, $\mathscr{P}$ are as above, then*

$$\gamma(Q \to d) = \gamma(Q^* \to d).$$

Proof. Recall that

$$\gamma(R \to d) = \frac{\sum |\{X \in \mathscr{P}(R) : X \text{ is } d\text{-deterministic}\}|}{|U|}.$$

If Y is a class $\mathscr{P}$, then

$$Z := \bigcup \{X \in \mathscr{P}(Q) : X \subseteq Y\}$$

contains exactly those elements of U which contribute to the Q-deterministic part of Y. Since Z is a class of Q^*, and every d-deterministic class of θ_{Q^*} has this form, the conclusion follows. □

6. Statistical justification

This section will show that usage of local binary rule reducts is useful to enhance the statistical strength of the prediction; we shall give an example in Section 7. In what follows, $\sum$ is the set of all permutations of U, and, as usual, H_0 denotes the null hypothesis.

Let σ be a permutation of U, and $p \subseteq \Omega$. We define new information functions $f_r^{\sigma(P)}$ by

$$f_r^{\sigma(P)}(x) := \begin{cases} f_r(\sigma(x)) & \text{if } r \in P, \\ f_r(x) & \text{otherwise.} \end{cases}$$

The resulting information system $\mathscr{I}_\sigma$ permutes the values within the P-rows according to σ, while leaving the Q-columns constant. We let $\gamma(Q \to \sigma(P))$ be the approximation quality of the prediction of $\sigma(P)$ by Q in $\mathscr{I}_\sigma$.

Given a rule $Q \to P$, we use the permutation distribution $\{\gamma(Q \to \sigma(P)): \sigma \in \sum\}$ to evaluate the strength of the prediction $Q \to P$. The value $p(\gamma(Q \to P) \mid H_0)$ measures the extremeness of the observed approximation quality and it is defined by

$$p(\gamma(Q \to P) \mid H_0) := \frac{|\{\gamma(Q \to \sigma(P)) \geqslant \gamma(Q \to P): \sigma \in \sum\}|}{|U|!} \tag{5}$$

This is the number of all permutations σ of U for which the approximation quality $\gamma(Q \to \sigma(P))$ is at least as large as the original one, normalized by the number of all permutations. If $p(\gamma(Q \to P) \mid H_0)$ is low, traditionally below 5%, then the rule $Q \to P$ is deemed significant.

The following shows that the filtration procedure does not decrease the statistical significance of a rule given below.

Proposition 4. *Let $Q \to P$, and Q^* be as defined just before Lemma 3. Then*

$$p(\gamma(Q \to P) \mid H_0) \geqslant p(\gamma(Q^* \to P) \mid H_0).$$

Proof. First, we observe that because attribute values are identified in the filtration process, each class of θ_{Q^*} is a union of classes of θ_Q. Thus, given any $R \subseteq \Omega$, the rule $Q \to R$ will have at least as many deterministic cases as $Q^* \to R$. It follows that $\gamma(Q \to R) \geqslant \gamma(Q^* \to R)$. Thus, for every $\sigma \in \Sigma$, if $\gamma(Q^* \to \sigma(P)) \geqslant \gamma(Q^* \to P)$, then we have

$$\gamma(Q \to \sigma(P)) \geqslant \gamma(Q^* \to \sigma(P)) \geqslant \gamma(Q^* \to P) = \gamma(Q \to P),$$

the latter by Lemma 3. Hence, the numerator of the right-hand side of Eq. (5) for $Q \to P$ is at least as large as that for $Q^* \to P$, whence the conclusion follows. □

7. An application with an empty core

Teghem and Charlet [15] search for premonitory factors of earthquakes by emphasizing gas geochemistry, measured over 155 days. The independent attributes are radon concentration and other measures of climatic factors, the decision attribute is the seismic activity, measured on the Richter scale, see Table 7.

The core with respect to the decision attribute is empty, and if we consult the literature on rough set analysis we get the advice that:

> ... Nondeterminism is particularly strong if the core knowledge is void. Hence nondeterminism introduces synonymy to the knowledge, which in some cases may be a drawback [10], p. 38.

Table 7
Earthquake system attributes

No	Attribute
0	1st Radon measure (location 1)
1	2nd Radon measure (location 1)
2	3rd Radon measure (location 2)
3	4th Radon measure (location 2)
4	5th Radon measure (location 2)
5	6th Radon measure (location 2)
6	7th Radon measure (location 2)
7	8th Radon measure (location 2)
8	Atmospheric pressure
9	Sun period
10	Air temperature
11	Relative humidity
12	Rainfall
13	Frost (measured at location 1)
14	Frost (measured at location 2)
15	Seismic activity

There are no tools within rough set analysis to proceed in this situation, and sometimes questionable procedures like counting of the appearance of attributes within rule reducts are applied to cope with this problem [15].

Because it is likely that the joint granularity of the attributes is too fine for a significant description of the decision attribute, the application of the data filtering procedure described in Section 5 suggests a way to improve the measurement of the attributes.

In the example of Table 7, one rule reduct of the decision attribute consists of the attributes RADON 11, RADON 21, RADON 32, RADON 42, RADON 52, RADON 62, where RADON XY means: The significance analysis tells us that the chance to obtain the rule randomly is close to 100%, and thus there is no evidence that the deterministic rule

RANDON 11, RANDON 21, RANDON 32, RANDON 42, RANDON 52, RANDON 62, $\Rightarrow$ SEISMIC ACTIVITY

is not due to chance.

The method of local reducts described in Section 5 shows that some variables can be recoded without loss of information by the following transformations:

- Radon 11 should be filtered by $\{3,4,5\} \to \{5\}$,
- Radon 21 should be filtered by $\{1,2\} \to \{1\}$,
- Radon 32 should be filtered by $\{1,5\} \to \{5\}$,
- Radon 62 should be filtered by $\{1,4\} \to \{1\}$.

The significance test of 400 simulations with the recoded data shows that the chance to obtain the rule at random is 0.25%, thus, we can assume that the recoded rule is significant.

8. Conclusions

We offer a technique of filtering attributes within rough set data analysis which has only marginal computational costs – in comparison to the standard procedures (like reduct searching) of RSDA – and which improves the strength of the results – i.e. the statistical significance of the rules–remarkably.

The technique can be applied as a cheap standard algorithm if the decision attributes P are fixed, and we have implemented the proposed procedure in GROBIAN, our engine for rough set data analysis [9]. Filtering leads to a conglomeration of equivalence classes which does not change the structure and the precision of the prediction.

If desired, the conglomeration process can be made to respect an ordinal structure of the attributes, by performing the union of equivalence classes only with classes which cannot be separated by "intermediate" classes. For example the elements 1 and 5 of Radon 32 would not be conglomerated, because they were separated by other measures (2, 3 and 4), whereas the conglomerations within Radon 11 and Radon 12 can be done, because the order structure of the value sets is not damaged by the procedure. More about rough set procedures which respect ordinal information of data values can be found in [16].

After having finished this paper we learned that our approach can be regarded as a special case of more sophisticated procedures, independently developed in [17], which can be called a "rough filtering" of attributes. We still feel that our approach is worthwhile since it easily implemented, it is less costly than the more advanced methods, and seems to be sufficient in many cases. Furthermore, the motivation for this paper, namely, to increase the statistical significance of rules by data filtering is an issue not discussed in [17].

References

[1] T.Y. Lin, N. Cercone (Eds.), Rough Sets and Data Mining, Kluwer Academic publishers, Dordrecht, 1997.

[2] Z. Pawlak, R. Słowiński, Rough set approach to multi-attribute decision analysis, ICS Research Report 36, Warsaw University of Technology, 1993.

[3] I. Düntsch, G. Gediga, Rough set dependency analysis in evaluation studies: an application in the study of repeated heart attacks, Inform. Res. Reports 10 (1995) 25–30.

[4] J. Rogner, M. Bartram, W. Hardinghaus, D. Lehr, A. Wirth, Depressiv getönte Krankheitsbewältigung bei Herzinfarktpatienten – Zusammenhänge mit dem längerfristigen Krankheitsverlauf und Veränderbarkeit durch eine Gruppentherapie auf indirekt-suggestiver

Grundlage, Coping. Verlaufs- und Therapiestudien chronischer Krankheit, G. Schübler, E. Leibing (Eds.), Hogrefe, Göttingen, 1994, pp. 95–109.
[5] G. Gigerenzer, Messung und Modellbildung in der Psychologie, Birkhäuser, Basel, 1981.
[6] I. Düntsch, G. Gediga, Statistical evaluation of rough set dependency analysis, International Journal of Human-Computer Studies 46 (1997) 589–604.
[7] E.S. Edgington, Randomization tests, Statistics: Textbooks and Monographs, vol. 31, Marcel Dekker, New York, 1987.
[8] B.F.J. Manly, Randomization and Monte Carlo Methods in Biology, Chapman and Hall, London, 1991.
[9] I. Düntsch, G. Gediga, The rough set engine GROBIAN, Proceedings of the 15th IMACS World Congress (Berlin), pp. 613–618.
[10] Z. Pawlak, Rough Sets: Theoretical Aspects of Reasoning about Data, Kluwer Academic Publishers, Dordrecht, 1991.
[11] L.R. Dice, Measure of the amount of ecologic association between species, Ecology 26 (1945) 297–302.
[12] P. Jaccard, Nouvelles recherches sur la distribution florale, Bulletin de la Society Vaudoise des Sciences Naturelless 44 (1908) 223–270.
[13] M. Novotný, Z. Pawlak, Algebric theory of independence, Fund. Inform. 14 (1991) 454–476.
[14] I. Düntsch, G. Gediga, Algebraic aspects of attribute dependencies in information systems, Fund. Inform. 29 (1997) 119–133.
[15] J. Teghem, J.-M. Charlet, Use of "rough sets" method to draw premonitory factors for earthquakes by emphasizing gas geochemistry: the case of a low seismic activity context in Belgium, Intelligent decision support: Handbook of Applications and Advances of Rough Set Theory, System Theory, Knowledge Engineering and Problem Solving, vol. 11, Kluwer Academic Publishers, Dordrecht, 1992, pp. 165–179.
[16] I. Düntsch, G. Gediga, Relation restricted prediction analysis (extended abstract), Proceedings of the 15th IMACS World Congress, Berlin, pp. 619–624.
[17] A. Skowron, L. Polkowski, Analytic morphology: Mathematical morphology of decision tables, Fund. Inform. 27 (7) (1996) 255–271.